Modelling the microbiome: deciphering the ecology of microbial communities

**Supervisors:** Dr Luke McNally (SoBS/IEB), Dr Jarrod Hadfield (SoBS/IEB), and Dr Roman Popat (Legible Data Limited)

**CASE partnership:** The CASE partner for this project will be Legible Data Limited, a data analytics company with a broad client base across the private and public sectors

**Primary supervisor contact email:** luke.mcnally@ed.ac.uk

**Project background:**
Modern sequencing technology has transformed our understanding of microbes and their ecology. Microbes exist in complex web of competing and cooperating species within their hosts, termed the ‘microbiome’. These communities have major effects on many aspects of host health including autoimmune disorders, diabetes, obesity, susceptibility to infections, and psychiatric disorders.

While many microbiome-sequencing projects have provided insight into the ecology of the microbes within us, a major impediment to advances in this field has been a lack of statistical methods to deal with the highly complex data that these studies generate. The central goal of this project will be to adapt statistical methods commonly used to study the ecology and evolution of plants and animals for application to microbiome data, and use these methods to provide new insight into microbial ecology and their effects on host health.

**Key research questions:**
How much variation in microbiome data is real biological variation and how much is experimental and sequencing noise? To what extent can we group microbes into “functional groups” with similar ecologies or is ecology species/strain specific? What is the evolutionary heritability of a microbe’s niche? What aspects of microbial ecology predict their effects on host health?

**Methodology:**
These questions will be addressed by developing a mixed effects modelling approach and novel machine learning approaches to model microbiome composition and its effects on host health while controlling for experimental and sequencing noise and the shared evolutionary history of species within the microbiome. These approaches will be applied to human, other animal and environmental microorganisms, depending on the student’s interests. The first 18 months of the project will focus on methods development, with the remaining 24 months focusing on application of these methods to microbiome datasets.

The approaches we will develop will be broadly applicable beyond microbial ecology and the student will have the opportunity to gain experience of their application to data from industry and the public sector during their training with our CASE partner, Legible Data Limited.

**Training:**
A comprehensive training programme will be provided comprising both specialist

*Microbes exist in complex multi-species communities within their hosts*
scientific training and generic transferable and professional skills. In addition, the student will receive specialist training in statistical modeling, data science, programming using R and Python, bioinformatics and microbiology. Training will be provided both by the academic supervisors (Dr McNally & Dr Hadfield) and by hands-on experience of application of data science in industry and public sector with Dr Roman Popat from our CASE partner Legible Data Limited.

Requirements:
This project would be suitable either for a student from a quantitative discipline (such as Mathematics, Statistics or Computer Science) wishing to gain knowledge and experience in ecology and microbiology, or a student from a biological discipline wishing to develop their skills in data science and statistics. Prior experience with programming and statistics would be a major advantage but is not essential.

Suggested Reading:

Please also see the web pages of the project supervisors and case partners

http://lukemcnally.wordpress.com/
http://jarrod.bio.ed.ac.uk/
https://www.legibledata.org/

Project summary: We are each home to a huge diversity of microbes. This project will develop new statistical methods to provide novel insight into microbial ecology and its consequences for host health.